GAUSSIAN MAXIMUM OF ENTROPY AND REVERSED LOG-SOBOLEV INEQUALITY

DJALIL CHAFAÏ

Abstract. The aim of this note is to connect a reversed form of the Gross logarithmic Sobolev inequality with the Gaussian maximum of Shannon's entropy power. There is thus a complete parallel with the well-known link between logarithmic Sobolev inequalities and their information theoretic counterparts. We moreover provide an elementary proof of the reversed Gross inequality via a two-point inequality and the Central Limit Theorem.

1. Shannon's entropy power and Gross's inequality

In the sequel, we denote by $\mathbf{Ent}_\mu(f)$ the entropy of a non-negative integrable function f with respect to a positive measure μ, defined by

$$\mathbf{Ent}_\mu(f) := \int f \log f \, d\mu - \int f d\mu \, \log \int f d\mu.$$

The Shannon entropy [15] of an n-variate random vector X with probability density function (pdf) f is given by

$$\mathbf{H}(X) := -\mathbf{Ent}_{\lambda_n}(f) = -\int f \log f \, dx,$$

where dx denotes the n-dimensional Lebesgue measure on $\mathbf{R}^n$. The Shannon entropy power [15] of X is then given by

$$\mathbf{N}(X) := \frac{1}{2\pi e} \exp\left(\frac{2}{n}\,\mathbf{H}(X)\right).$$

It is well-known (cf. [15, 8]) that Gaussians saturates this entropy at fixed covariance. Namely, for any n-variate random vector X with covariance matrix $\mathbf{K}(X)$, one have

$$\mathbf{N}(X) \le |\mathbf{K}(X)|^{1/n}, \tag{1}$$

and $|\mathbf{K}|^{1/n}$ is the entropy power of the n-dimensional Gaussian with covariance $\mathbf{K}$.

The logarithmic Sobolev inequality of Gross [11] expresses that for any non-negative smooth function $f : \mathbf{R}^n \to \mathbf{R}^+$

$$2\,\mathbf{Ent}_{\gamma_n}(f) \le \mathbf{E}_{\gamma_n}\left(\frac{|\nabla f|^2}{f}\right), \tag{2}$$

where $\mathbf{E}_{\gamma_n}$ denotes the expectation with respect to γ_n, $|\cdot|$ the Euclidean norm and γ_n the n-dimensional standard Gaussian given by

$$d\gamma_n(x) := (2\pi)^{-\frac{n}{2}} e^{-\frac{|x|^2}{2}}\, dx.$$

Inequality (2) is sharp and the equality is achieved for f of the form $\exp(a\cdot)$.

By performing a change of function and an optimization, Beckner showed [4] (see also [7]) that (2) is equivalent to the following "Euclidean" logarithmic Sobolev inequality, for any pdf g

$$\mathbf{Ent}_{\lambda_n}(g) \leq \frac{n}{2}\log\left[\frac{1}{2\pi e n}\int\frac{|\nabla g|^2}{g}\,dx\right], \tag{3}$$

where λ_n is the n-dimensional Lebesgue measure on $\mathbf{R}^n$. Therefore, for any n-variate random vector X (with pdf g), we have

$$\mathbf{N}(X)\,\mathrm{J}(X) \geq n. \tag{4}$$

This inequality can be obtained by many methods. The most classical ones are via Shannon's entropy power inequality together with DeBruijn identity, or via Stam's super-additivity of the Fisher information (cf. [16, 7, 10, 1]). Moreover, Dembo showed in [9] that (4) is equivalent to

$$\mathbf{N}(X)\,|\mathbf{J}(X)|^{1/n} \geq 1, \tag{5}$$

where $\mathbf{J}(X)$ is the Fisher information matrix of X defined by

$$\mathbf{J}(X) := \int\nabla\log g\cdot\nabla\log g^\top\,g\,dx,$$

and we have $\mathrm{J}(X) = \mathrm{Tr}\,\mathbf{J}(X)$. To deduce (5) from (4), apply (4) to the random vector $X = \mathbf{K}(Y)^{-1/2}\,Y$. Conversely, use the arithmetic-geometric means inequality

$$(a_1\cdots a_n)^{\frac{1}{n}} \leq \frac{a + \cdots + a_n}{n} \tag{6}$$

on the spectrum of the non-negative symmetric matrix $\mathbf{J}(X)$.

2. Reversed Gross's logarithmic Sobolev inequality

The Gross logarithmic Sobolev inequality (2) admits a reversed form which states that for any positive smooth function $f : \mathbf{R}^n \to \mathbf{R}^+$

$$\frac{|\mathbf{E}_{\gamma_n}(\nabla f)|^2}{\mathbf{E}_{\gamma_n}(f)} \leq 2\,\mathbf{Ent}_{\gamma_n}(f)\,. \tag{7}$$

Here again, the 2 constant is optimal and the equality is achieved for f of the form $\exp(a\cdot)$. Alike for (2), one can show by a change of function and an optimization that the reverse form (7) is equivalent to the following inequality, for any pdf g

$$-\mathbf{Ent}_{\lambda_n}(g) \leq \frac{n}{2}\log\left[2\pi e\frac{\mathrm{Tr}\,\mathbf{K}(g)}{n}\right], \tag{8}$$

where $\mathbf{K}(g)$ is the covariance matrix of the pdf g. Hence, we have for any n-variate random vector X with pdf

$$\mathbf{N}(X) \leq \frac{\mathrm{Tr}\,\mathbf{K}(X)}{n}, \tag{9}$$

where $\mathbf{K}(X)$ denotes the covariance matrix of X. This inequality is optimal and is achieved by Gaussians X. Moreover, as we will show, inequality (9) is equivalent to (1).

Summarizing, we obtain the following statement

Theorem 1. *The following assertions are true and equivalent*

(i) *For any smooth* $f : \mathbf{R}^n \to \mathbf{R}^+$,

$$|\mathbf{E}_{\gamma_n}(\nabla f)|^2 \leq 2\,\mathbf{Ent}_{\gamma_n}(f)\,\mathbf{E}_{\gamma_n}(f)\,.$$

(ii) *For any smooth* $g : \mathbf{R}^n \to \mathbf{R}^+$,

$$-\mathbf{Ent}_{\lambda_n}(g) \leq \frac{n}{2}\log\left[\frac{2\pi e}{n}\operatorname{Tr}\mathbf{K}(g)\right].$$

(iii) *For any n-variate random vector X with smooth pdf,*

$$n\mathbf{N}(X) \leq \operatorname{Tr}\mathbf{K}(X)\,.$$

(iv) *For any n-variate random vector X with smooth pdf,*

$$\mathbf{N}(X) \leq |\mathbf{K}(X)|^{1/n}.$$

Therefore, there is a complete parallel between the equivalence between (2), (3), (4), (5) in one hand and the equivalence between (7), (8), (9), (1) in the other hand.

3. Sketches of proofs

In this section, we present first two proofs of (7), then we explain how to deduce (8) from (7) and (1) from (9) and vice versa.

The most natural way to establish (7) is to start from a two-point. inequality, just like Gross does for the logarithmic Sobolev inequality (2) in [11]. Namely, if we denote by β the symmetric Bernoulli measure on $\{-1,+1\}$, one can show easily that for any non-negative function $f : \{-1,+1\} \to \mathbf{R}^+$,

$$(f(+1) - f(-1))^2 \leq \frac{1}{8}\,\mathbf{Ent}_\beta(f)\,\mathbf{E}_\beta(f)\,. \tag{10}$$

This inequality is nothing else but the Csiszár-Kullback inequality for β (see [14]). Actually, the optimal constant for the Bernoulli measure of parameter $p = 1-q$ is $p^2q^2(\log q - \log p)/(q-p)$, which resembles the optimal constant for the logarithmic Sobolev inequality (see for example [1]), but here again, only the symmetric case gives the optimal constant of (7).

The next step is to establish the following chain rule formula for $\mathbf{Ent}$, which generalizes the classical chain rule formula (cf. [15, 8]) for $\mathbf{H}$

Proposition 2. *For any positive measures μ_i and their product μ on the product space, and for any bounded real valued measurable function f on the product space, we have*

$$\mathbf{Ent}_\mu(f) \geq \sum_{i=1}^n \mathbf{Ent}_{\mu_i}\left(\mathbf{E}_{\mu_{\setminus i}}(f)\right), \tag{11}$$

where $\mu_{\setminus i}$ denotes the product of the measures μ_j with $j \neq i$.

Finally, inequality (7) can then be recovered by the use of the Central Limit Theorem and integration by parts (in both discrete and Gaussian forms). This concludes the first proof of (7).

Actually, inequality (7) can be recovered by a simple semi-group argument, just like for the logarithmic Sobolev inequality (2) (cf. [13]). Namely, consider the heat semi-group $(\mathbf{P}_t)_{t\geq 0}$ on $\mathbf{R}^n$, acting on a bounded continuous function $f : \mathbf{R}^n \to \mathbf{R}$ as follows

$$\mathbf{P}_t(f)\,(x) := \int_{\mathbf{R}^n} f(x + \sqrt{t}\,y)\,d\gamma_n(y).$$

Notice that for any smooth function f, $\nabla \mathbf{P}_t(f) = \mathbf{P}_t(\nabla f)$ and

$$\partial_t \mathbf{P}_t(f) = \frac{1}{2}\, \Delta \mathbf{P}_t(f) = \frac{1}{2}\, \mathbf{P}_t(\Delta f)\,.$$

Now, for any smooth positive bounded function $f : \mathbf{R}^n \to \mathbf{R}^+$, any $t \geq 0$ and any x, we can write, by performing an integration by parts and omitting the x variable

$$\begin{aligned}\mathbf{P}_t(f \log f) - \mathbf{P}_t(f) \log \mathbf{P}_t(f) &= \int_0^t \partial_s [\mathbf{P}_s(\mathbf{P}_{t-s}(f) \log \mathbf{P}_{t-s}(f))]\, ds \\ &= \frac{1}{2} \int_0^t \mathbf{P}_s\left(\frac{|\nabla \mathbf{P}_{t-s}(f)|^2}{\mathbf{P}_{t-s}(f)}\right) ds.\end{aligned}$$

But by Cauchy-Schwarz inequality we get

$$\mathbf{P}_s\left(\frac{|\nabla \mathbf{P}_{t-s}(f)|^2}{\mathbf{P}_{t-s}(f)}\right) = \mathbf{P}_s\left(\frac{|\mathbf{P}_{t-s}(\nabla f)|^2}{\mathbf{P}_{t-s}(f)}\right) \geq \frac{|\mathbf{P}_s(\mathbf{P}_{t-s}(\nabla f))|^2}{\mathbf{P}_s(\mathbf{P}_{t-s}(f))},$$

which gives

$$\mathbf{P}_t(f \log f) - \mathbf{P}_t(f) \log \mathbf{P}_t(f) \geq \frac{t}{2} \frac{|\mathbf{P}_t(\nabla f)|^2}{\mathbf{P}_t(f)}.$$

Finally, inequality (7) follows by taking $(t, x) = (1, 0)$. Notice that this method gives also the logarithmic Sobolev inequality (2). Namely, by Cauchy-Schwarz inequality

$$|\mathbf{P}_{t-s}(\nabla f)|^2 \leq \mathbf{P}_{t-s}(|\nabla f|)^2 \leq \mathbf{P}_{t-s}(f)\, \mathbf{P}_{t-s}\left(\frac{|\nabla f|^2}{f}\right),$$

therefore, we obtain

$$\begin{aligned}\mathbf{P}_t(f \log f) - \mathbf{P}_t(f) \log \mathbf{P}_t(f) &= \frac{1}{2} \int_0^t \mathbf{P}_s\left(\frac{|\mathbf{P}_{t-s}(\nabla f)|^2}{\mathbf{P}_{t-s}(f)}\right) ds \\ &\leq \frac{1}{2} \int_0^t \mathbf{P}_s\left(\mathbf{P}_{t-s}\left(\frac{|\nabla f|^2}{f}\right)\right) ds \\ &= \frac{t}{2}\, \mathbf{P}_t\left(\frac{|\nabla f|^2}{f}\right),\end{aligned}$$

which gives (2) by taking here again $(t, x) = (1, 0)$.

To deduce (8) from (7), just apply (7) to

$$f(x) = h(x)\, (2\pi)^{\frac{n}{2}}\, e^{\frac{|x|^2}{2}}$$

where h is a compactly supported smooth pdf. One then gets

$$\left|\int x\, h\, dx + \int \nabla h\, dx\right|^2 \leq 2 \int h \log h\, dx + \int |x|^2\, h\, dx + n \log(2\pi).$$

But we have $\int \nabla h\, dx = 0$. Therefore, by denoting $\mathbf{K}(h)$ the covariance matrix associated with the pdf h, one gets

$$-\mathbf{Ent}_{\lambda_n}(h) \leq \frac{1}{2} \operatorname{Tr} \mathbf{K}(h) + \frac{n}{2} \log(2\pi),$$

which remains true for any smooth pdf h. Finally, by performing the change of function $h = \alpha g(\alpha\cdot)$ and optimizing in α, one obtains

$$-\mathbf{Ent}_{\lambda_n}(h) \leq \frac{n}{2} \log \left[2\pi e \frac{\operatorname{Tr} \mathbf{K}(h)}{n} \right]$$

which is nothing else than (9). Conversely, it is easy to see that we can recover (7) for any pdf f by approximating f by compactly supported probability density functions.

The equivalence between (9) and (1) is obtained as for the equivalence between (4) and (5). Namely, to deduce (1) from (9), apply (9) to the random vector $X = \mathbf{K}(Y)^{-1/2} Y$. Conversely, use the arithmetic-geometric means inequality (6) on the spectrum of the non-negative symmetric matrix $\mathbf{K}(X)$.

4. Remarks

It is well-know that the logarithmic Sobolev inequality (2) is a consequence of the Gaussian isoperimetric inequality [13]. In contrast, it is shown in [3] that the reversed form (7) is equivalent to a translation property [6]. Namely, for any smooth function $f : \mathbf{R}^n \to [0,1]$

$$|\mathbf{E}_{\gamma_n}(\nabla f)| \leq \mathbf{I}(\mathbf{E}_{\gamma_n}(f)), \tag{12}$$

where $\mathbf{I}$ is the Gaussian isoperimetric function given by $\mathbf{I} := \Phi' \circ \Phi^{-1}$, where Φ is the Gaussian distribution function given by $\Phi(\cdot) := \gamma((-\infty, \cdot])$. Bobkov's inequality (12) expresses that among all measurable sets with fixed Gaussian measure, half spaces have minimum barycenter [3].

Acknowledgments. The author would like to thanks Michel Ledoux and Cécile Ané for their encouragements and helpful comments.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. Sur les inégalités de Sobolev logarithmiques. to appear in "Panoramas et Synthèses, Société Mathématique de France, 2001.

[2] D. Bakry, D. Concordet, and M. Ledoux. Optimal heat kernel bounds under logarithmic Sobolev inequalities. *ESAIM Probab. Statist.*, 1:391–407 (electronic), 1997.

[3] F. Barthe, D. Cordero-Erausquin, and M. Fradelizi. Shift inequalities of gaussian type and norms of barycenters. preprint, september 1999.

[4] W. Beckner. Geometric asymptotics and the logarithmic Sobolev inequality. *Forum Math.*, 11(1):105–137, 1999.

[5] N. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Information Theory*, IT-11:267–271, 1965.

[6] S. Bobkov. The size of singular component and shift inequalities. *Ann. Probab.*, 27:416–431, 1999.

[7] E. Carlen. Super-additivity of Fisher's information and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 101(1):194–211, 1991.

[8] T. Cover and J. Thomas. *Elements of information theory.* John Wiley & Sons Inc., New York, 1991. A Wiley-Interscience Publication.

[9] A. Dembo. Information inequalities and uncertainty principles. In *Tech. Rep., Dept. of Statist.* Stanford Univ., 1990.

[10] A. Dembo, T. Cover, and J. Thomas. Information-theoretic inequalities. *IEEE Trans. Inform. Theory*, 37(6):1501–1518, 1991.

[11] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[12] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII*, Lecture Notes in Math., pages 120–216. Springer, Berlin, 1999.

[13] M. Ledoux. The geometry of Markov Diffusion Generators. *Ann. Fac. Sci. Toulouse Math.*, IX(2):305–366, 2000.

[14] M. S. Pinsker. *Information and information stability of random variables and processes.* Holden-Day Inc., San Francisco, Calif., 1964. Translated by Amiel Feinstein.

[15] C. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[16] A. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, 1959.

E-mail address: chafai@cict.fr

Laboratoire de Statistique et Probabilités,UMR CNRS 55830, Université Paul Sabatier, F-31062 CEDEX, Toulouse, France.